Congratulations! You passed!
TO PASS 80% or higher

Keep Learning

GRADE
100%

Understanding Features

TOTAL POINTS 3

1. "Binning" features means: **1 / 1 point**

- ○ Assigning images to labelled bins or categories
- ◉ Turning continuous values into numbered, discrete sets
- ○ Getting rid of invalid or incomplete features
- ○ Defining relevant age ranges
- ○ Categorizing each feature as valid or invalid

Correct
Yes! Rather than having real values, we define "bins" and group all numbers within those defined ranges into one category.

2. Should you include every possible feature when preparing data for your learning algorithm? **1 / 1 point**

- ◉ No, because it makes the problem space too big.
- ○ No, because the learning algorithm will be distracted by bad signals
- ○ Yes, to avoid the curse of dimensionality
- ○ Yes, because more data is always better
- ○ Yes, because humans and machine learn from different signals

Correct
Correct! The "curse of dimensionality" describes how every new feature adds a new dimension to the hypothesis space the learning algorithm is searching in, making it exponentially larger. See the video on "how many features" for more details.

3. How do you know if a feature should be included for a prediction task? Select all that apply. **1 / 1 point**

- ☐ Because it's complete
- ☑ Because a domain expert said so

Correct
True! As always, it is import to listen to the human experts in your area. It may or may not end up being significant to the learning algorithm, but if an expert says it's useful it's important to consider. See the video on useful/useless features for more detail.

- ☑ Because the machine learning algorithm said so

Correct
True! If we're talking strictly about prediction, we can't always tell what is correlated with the prediction we care about. So if the machine learning algorithm picks up on a feature, it is at least correlated with the correct answer. See the video on useful/useless features for more detail.

- ☐ Because it was found through unsupervised learning
- ☐ Because the meaning varies over time
- ☐ Because it's easy to explain the importance
- ☑ Because it's highly correlated with the label

Correct
Correct! If we happen to know a feature has high correlation with the label, it is almost certainly an important feature to include.

- ☐ Because it varies a lot over the dataset
- ☐ Because it's highly correlated with other features